DOCUMENT RESUME

ED 449 165 TM 032 243

AUTHOR Ediger, Marlow
TITLE Portfolios versus State Mandated Testing.
PUB DATE 2000-00-00
NOTE 7p.
PUB TYPE Opinion Papers (120)
EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Academic Achievement; Elementary Secondary Education; Evaluation Methods; *Portfolio Assessment; Portfolios (Background Materials); Reliability; *Standardized Tests; *State Programs; *Student Evaluation; *Testing Programs

ABSTRACT

The use of portfolios in assessing student achievement is contrasted with the use of standardized tests. Portfolios are considered more subjective than scores from standardized tests, and they present different issues in assessing and reporting student progress. Interscorer reliability may be a problem because of a lack of agreement about the results of each portfolio. State mandated tests with multiple choice items may be either standardized (norm-referenced) or criterion-referenced tests. There are many problems with the use of standardized tests, including a high degree of mismatch between what is tested and what may have been taught in the classroom. Much of what is tested has been learned outside of school, and students from disadvantaged environments may be penalized in test taking. Such problems should not eliminate the use of testing to determine student achievement, but improvements must be made in the use of standardized tests. As the assessment of student problems takes on an increased role in society, better means of evaluation must be discovered and implemented. (SLD)

Reproductions supplied by EDRS are the best that can be made
from the original document.

Portfolios Versus State Mandated Testing

Marlow Ediger

ED 449 165

TM032243

F. Ediger

U.S. DEPARTMENT OF EDUCATION
Office of Educational Research and Improvement
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)
This document has been reproduced as received from the person or organization originating it.
Minor changes have been made to improve reproduction quality.
Points of view or opinions stated in this document do not necessarily represent official OERI position or policy.

BEST COPY AVAILABLE

PORTFOLIOS VERSUS STATE MANDATED TESTING

Portfolios and their use in assessing student achievement are considered to more subjective as compared to scores from standardized tests. Portfolios do not contain the objective numerals such as percentiles, stanines, standard deviations, quartile deviations, and grade equivalency as do state mandated test results. Even though tests do contain numerical results from student testing, they still possess the human element with

1. writing and editing of test items.
2. doing pilot studies and making revisions to the test.
3. developing scoring keys to use in the assessment process.
4. emphasizing item analysis results from the tests to arrange test items sequentially as well as to eliminate and modify the original items on the test.
5. finalizing the test, if standardized, to spread students out from high to low on the test results. Certain test items are then omitted from pilot study results. Thus, good test items are those in which a high scorer on the total test responds correctly to any single multiple choice item. Conversely, a bad test item in the pilot study is one which is responded to correctly by those who score low on the total test.

All tests have standard errors of measurement or weaknesses in reliability. A further problem is validity. It truly is difficult to write test items covering that which has been taught in the many classrooms within a state or within the nation.

Since portfolios are less concerned with the concepts of objectivity and numerical results, they have different difficulties in assessment and reporting student progress A rubric may be carefully developed and used to appraise portfolios, but the five point rubric standards, general in nature, may be difficult to implement when assessing each portfolio. Assessors of portfolios might well disagree on the meaning of each of the standards on the five point scale and thus make for great variation in student's rubric score. Interscorer reliability then becomes a problem due to a lack of agreement as to the results of each portfolio from the many being assessed (Ediger, 1999, 233-240).

Portfolio Development

The student with teacher guidance develops his/her very own portfolio. Contents in the portfolio are to show progress and achievement of the student. A representative sampling of products and processes are to go into the portfolio. Each item chosen represents a choice whereby objectivity is not involved. Thus, items such as the following may be chosen to become a part of the student's portfolio:

1. snapshots of construction items, art work, and dioramas made

in ongoing units of study.

2. cassette recordings of student oral communication experiences. These include book reports, talks, debates, public speaking, and oral reading, among others, within units of study.

3. a video-tape of committee work participation.

4. written products including narrative, expository, creative, poetry, and prose.

5. teacher assessment of the student in essay form as well as student self evaluation, using a five point rating scale, on carefully spelled out criteria.

By reflecting upon his/her portfolio contents, the student realizes what has been learned, what is left to learn, and what needs the most attention to realize optimal achievement. The parent(s) may also look at sequential entrees to realize how well the student is achieving and what needs more attention. The portfolio might well be observed and discussed in a parent/teacher conference. A carefully devised rubric might be used to assess each portfolio by experienced assessors. Hopefully, Interscorer or interrater reliability will be in the offing. The results from rubric assessment will yield a numeral which will be quite subjective as compared to a very specific agreed upon scoring key providing numerical result from a state mandated test. In state mandated testing, the same key is used to score all multiple choice test results. With machine scoring, many tests can be scored in a short time with a printout to show how well each student has done (See Murphy, 1997, 81).

State Mandated Testing

State mandated tests with multiple choice items may be either standardized or a criterion referenced test (CRT). Standardized tests are generally developed by a commercial company in which the items have been tried out in pilot studies. With pilot studies on a random sampling of students, bad test items may be eliminated or modified. Bad test items lack clarity in meaning and are poorly written. There is an additional screen used in accepting/rejecting test items from pilot study results. Thus, a test item to be acceptable must be answered correctly by those receiving highest scores on the total test. a negative test item is answered correctly by those having the total lowest test score on the test. Conversely, a test item is negative if missed by those scoring highest on the total test of multiple choice test items. Also, a good test item is one in which the lowest scorers on the total test responded incorrectly to any one test items. The goal is to spread students out from the 99th to the first percentile. Most classrooms will not have this large a spread of scores in terms of student achievement. However for the total

number of students taking the pilot study test, the final results will amount to a general bell shaped curve (See Ediger, 2000, 155-161).

The norms of the standardized test provide information for placing the local student's results on the continuum for percentiles, grade equivalents, standard deviations, quartile deviations, and/or stanines.

A critical evaluation of standardized testing was given by Wesson (Education Week, 2000) in which he listed the following criticisms:

1. those who know he least about learning and child development are the strongest advocates of their use to measure student performance in schools.
2. parents whose children are winners on standardized tests push for broader use of these tests.
3. these tests were never meant to measure educational quality nor teaching excellence.
4. the education level and occupation of parents, economic advantages, and location of the schools attended are important factors in how well a student does on a standardized test.
5. standardized tests were never meant to measure accountability of teachers.
6. these tests are designed to provide for variation in achievement among students so that some will be "left behind." Thus, 50% of students taking the test will be below the mean and 50% above the mean. Writers, from pilot studies, choose multiple choice items whereby 50% responded correctly to any one numbered item on the standardized test. If, for example 100% of the students responded correctly to a test item, there would be no spread of scores and no variation among test takers.
7. there is a high mismatch between what is tested and what has been taught. Much of what is tested has been learned by students outside of school. Students from poor homes are penalized in test taking.
8. students with limited English proficiency (LEP) will tend to do poorly on standardized tests.
9. filling in the bubble on an answer sheet to test items reduces the meaning of superior education.
10. standardized tests do not measure highly valuable unquantifiable traits such as perseverance, intuition, adaptability, responsibility, sensitivity, empathy, self-control, motivation, effective communication skills, friendliness, honesty, kindness, commitment, loyalty, emotional maturity, inventiveness, cooperativeness, and trustworthiness.

What can be measured is an important criteria for what goes into a standardized test. Quantifiable test results are then wanted. Facts and factual information can easily be measured to ascertain student achievement, but higher levels of cognition, including critical and creative thinking as well as problem solving, presents tremendous

problems when writing multiple choice test items. Human qualities such as kindness, present even a further difficulty in testing. Thus, how much kindness does any one person possess? Then too, there are no opportunities to show achievement in any way for oral and written communication.

Do these comments eliminate the use of testing to ascertain student achievement. The answer would be, "no." What has been listed as criticisms need to be analyzed and improvements made for the further use of tests to determine student achievement. Also, multiple assessments need to be used. Multiple Intelligences Theory (see Gardner, 1993) indicates to the teacher that there are numerous ways for students to reveal achievement and progress. The classroom teacher, however, may have little or no input into how students are to be assessed. States do determine what will be in terms of evaluative procedures. There is some room though for teachers to decide how to assess student achievement. This "room" is minimized much with high stakes testing. Why? Teachers feel the pressure to drill students on what might be on a standardized test. Weeks and even months may be spent here on rote learning of facts, the lowest level of cognition. Teacher then do not want to be scolded for low student achievement on standardized tests. Nor, do they want to worry about being dismissed because of poor student test results.

Problems in Student Assessment

There are definite problems which need to be ironed out when scoring/evaluating how well a student is doing in school or in general. These problems include the following:

1. how much of student achievement can be indicated from testing with a score or a single numeral, such as a percentile, rather than observing actual daily products/processes of learners in a portfolio?
2. how much stress should be placed upon evaluating personality traits, such as growth in perseverance on a project/activity in school and in life?
3. how much input should come from external sources such as test writers in determining school achievement of learners?
4. how much input should come from the classroom teacher and/or the learner himself/herself in evaluating achievement?
5. how much focus should be placed upon sticking to the academics in teaching as compared to other worthwhile endeavors, such as cooperation and utilitarian endeavors? Academic learnings, in most cases, may be easier to assess using testing procedures. Cooperation needs to be inferred from observing student behavior and does not lend itself to paper/pencil testing nor from assigning an agreed upon numeral by assessors.

6. how do standardized means of assessing compare with rubric use, for example, in evaluating the quality of written/oral work?

7. how accurate is computerized machine scoring of large numbers of state mandated tests? There have been bad computer glitches in test results from students. Might appropriate validity and reliability come about from quality rubrics to assess products and processes?

8. how do the attitudes of the lay public compare using single test scores to appraise student achievement versus more inclusive data from daily student work in portfolios?

9. how can test scores and portfolio assessments be used more effectively to improve the curriculum for all learners?

10. how cost effective is paying for machine scoring of tests as compared to human appraisal of student portfolios?

As assessment of student progress and achievement take on increased interest and purpose in society, better means of evaluation need to be discovered and implemented. This is a challenge for educators and interested persons in the societal arena.

References

Ediger, Marlow (1999), "Appraising Learner Progress in the Social Studies," College Student Journal, 33 (2) 233-240.

Ediger, Marlow (2000), "The Principal and Evaluation of Student Achievement," The Journal of Instructional Psychology, 27 (3), 155-161.

Gardner, Howard (1993), Multiple Intelligences: Theory into Practice. New York: Basic Books.

Murphy, Sandra (1997), "Designing Portfolio Assessment, Programs to Enhance Learning," Clearing House, 74 (2), 81.

Wesson, Kenneth A. (November 22, 2000), "The Volvo Effect," Education Week, pp. 34, 36, 37.